大数据时代数据新闻可视化的发展探讨

摘　要：随着智能设备的普及和互联网技术的飞速发展，大数据这一概念应运而生。在技术的强力支撑下，新闻传媒业从琐碎冗杂的结构化、非结构化数据中，挖掘有意义的信息，将其整合分析，发现其中的深度价值，并由此产生一种跨学科、跨领域的新型新闻生产方式——数据新闻。数据新闻丰富了调查方法和叙事手段，增强了网络新闻的真实性、客观性、可读性，但也面临威胁和挑战。本文试图从多个层面进行分析，探索数据新闻可视化发展过程中遇到的问题，并提出针对性建议，旨为在数据新闻的发展尽绵薄之力。

关键词：大数据；数据新闻；可视化
中图分类号：G22　　文献标识码：A
文章编号：1671-0134（2017）08-120-04　　DOI：10.19483/j.cnki.11-4653/n.2017.08.047

■文／肖卓明　罗琪元　吴　娴

1. 数据新闻可视化概述

数据可视化伴随着统计学的出现而产生，因计算机的诞生而向前大步发展。随着 IT 技术、互联网、移动互联网技术的蓬勃发展，人类的一言一行都被智能记录、快速处理并且表达出来，人类通过互联网、智能软硬件产生了大量结构化、非结构化数据，大数据的概念应运而生。随着云计算、分布式处理技术、存储技术和感知技术的普及，大数据在各个行业得到了充分的应用探索。大数据的概念也在悄然改变新闻行业，其中最具代表性的成果就是数据新闻的诞生与发展。

数据新闻实践始于国外优秀媒体，如英国《卫报》、BBC 广播电台、《华尔街日报》《纽约时报》等在 21 世纪初就开始进行数据新闻领域的探索和实践。通过网络协作的方式，一些媒体从业者合力编写了著名的《数据新闻手册》，具有极高的实用性和实践价值。随着数据新闻的发展，越来越多的机构、新闻从业者、自由撰稿人以及一些爱好者纷纷创办专门的网站展示其数据新闻作品，内容涉及众多领域。

相比起主流媒体，数据新闻起步较晚。随着数据新闻生产理念的导入和大数据时代新闻数据化的自然趋势，如今，这一领域已经成为新媒体转型的必争之地，包括网易“数读”、搜狐“数字之道”、财经网“数字说”、新浪“图解天下”等，都致力于利用数据来呈现事实，提供轻量化的阅读体验。依托于自媒体平台传播的便利性，加上 H5 技术的发展，传统媒体最近几年在数据新闻可视化上也是下足了功夫，微信公众号、头条号上都不时会出现惊艳的数据可视化报道。优秀的可视化新闻可以通过自媒体平台得到快速传播，使传统媒体的影响力得到了更大范围的延伸。而《新京报》《南方都市报》《东方早报》《壹读》《南风窗》等报刊，甚至在纸质媒体上设置了固定的数据新闻版面。可以说，数据新闻可视化已经由边缘走向舞台中央。

数据新闻有科学性的一面，大数据时代让传播学和社会学变得越来越科学。如何从海量的数据信息挖掘出有新闻价值的部分，分析其中的结构和变化规律，经过加工整合，使用诸如图表、信息图等方式展示数据，让读者更加容易理解信息本身，并且了解信息的重要性和相关性，这是可视化的意义所在。当下的新闻市场中，对事件进行单一维度的解读，已经不太能够满足大众的需求，他们更倾向于对数据的多维分析，这就要求新闻从业者需要多角度、立体化解读新闻事件，深挖数据背后的关系链，这也是数据新闻可视化为什么能够得到推广应用的原因。

数据新闻可视化丰富了调查方法和叙事手段，增强了新闻的真实性和客观性，加速了传统媒体行业角色的转换，以预测趋势和把握舆论等方式推动社会发展。因此，数据新闻可视化的运用越来越广泛。但是，在发展的过程中，也遇到种种问题。

2. 数据新闻可视化的实践问题

早期对数据可视化与新闻报道相结合的尝试和实践，开启了数据新闻可视化的先河，发展至今效果显著。但在数据新闻可视化发展过程中，也面临着极大的挑战，如数据的模糊性、公开透明度、可能诱发的犯罪或者社会危机、采编队伍数据素养参差不齐，以及数据可视化工具普及程度不高等。当前在数据新闻可视化方面存在以下问题。

2.1 数据资源开放有限，获取的数据准确性不高

数据新闻的数据来源多样，如来自单位、企业公开发布的信息、国际性的数据调研机构调研发布的信息、社交媒体上用户行为产生的数据、媒体网站积累的数据等，但很多数据并未有合法渠道开放给大众。近年来，我国不断推动信息公开，但是大部分公开的数据还是报告和报表，而非原始数据，因此，无法对源数据进行更深入的分析、加工和挖掘。此外，社会各行业对数据价值的保密意识逐渐提高，在主动公开的信息上具有选择性。数据的公开程度低，是数据新闻发展的一大障碍。数据的开放性和共享性是数据新闻发展壮

大过程中亟待解决的问题。

虽然大数据具有强大的数据统计和数据整合功能，但凡是人为统计的数据，或多或少存在一定的模糊性。数据新闻作品基本上源自于已有的调查研究结果，是对历史数据进行的二次价值挖掘。倘若大规模的数据失真，在此基础上进行的数据分析进而得出的结果将大幅降低其指导价值，甚至很有可能会造成难以预估的后果。数据的“真”决定了新闻的价值所在，也是对读者担负的一种责任。这就需要在数据收集时，对其进行一定的甄别，尽量避免由于低级错误而导致数据失真的产生。

2.2 个人隐私存在安全隐患，数据垄断导致社会不安

大数据为监测和预示人们的生活提供了极大的方便，然而，个人隐私也随之暴露在无形的第三只眼之下。无论是电子商务、搜索引擎还是社交网站等互联网服务商都对用户行为进行挖掘和分析，获得商业利益，这一过程不可避免地威胁到普通人的隐私。若数据落入不良商家手中，必然会引发一场狂热的商业风暴。针对消费者的行为偏好和出行轨迹，商家可能会不择手段地推销自己的产品，甚至于采用一些极端的方式等。隐私权无法得到保障，在缺乏有效的伦理机制下，很容易引发社会问题，造成对公平、自由、平等和尊严等人性价值的践踏。

与数据产业发展相伴而生的是各种冲突和矛盾，数据垄断就是其中之一。数据垄断是指重要数据被控制在少数人手里，并被不合理地分配和享用。在大数据时代下，商家可以根据数据对商品进行价格控制或垄断，从而造成经济上的垄断。如若恶意转手甩卖，很有可能增加社会的不稳定因素。

2.3 数据新闻可视化应用程度不深

虽然数据新闻可视化在新媒体、传统媒体的推动下已经获得了一定程度的发展，并且也出现了一些优秀的作品，但是从整体上看，数据新闻可视化在国内的应用程度并不深入，主要表现在以下几个方面。

2.3.1 可视化形式单一，应用缺乏深度性

由于静态信息图相比于各类动态图形制作，成本与速度有一定的优势，所以国内数据新闻作品大多采用静态的形式，利用折线图、饼图、柱状图等信息图表辅助文字信息以表达。而少数动态的交互作品，也主要以时间轴、幻灯的形式来展现，视觉要素表现类似，新闻制作模式化。从整体来看识别度不高，动感不足，缺乏创新性，与读者的互动性有待提高。

数据本身并不具备价值，只有实现对数据的分析解读，以及深层次的数据挖掘，才能赋予数据新的价值和功能。但是目前，国内大多数的数据新闻仅仅充当吸引读者阅读的诱因，阐述了某个新现象或者理清某种结构的信息图，并不能完整地讲述新闻事件的始末，有机地融合在其他稿件中，对数据新闻内含的传达方面没有与深刻性完美契合。

2.3.2 预测性可视化新闻匮乏

通过特定的计算机算法可以对大量的数据进行自动分析，挖掘数据中存在的隐性关系和模式。通过收集的数据，合理地预测出事物未来的发展动向和潜在关系。《大数据时代》的作者维克托·迈尔·舍恩伯格曾说：大数据的核心就是预测，不是要教机器像人一样思考，而是要把数字计算运用到海量的数据上，来预测事情发生的可能性。但是当前的数据新闻中，预测性新闻占比非常之小，更多的是一些解释性的说明图，并没有做到充分挖掘数据价值，研究数据背后的关联。通过已有数据，预测出事物未来的发展动向和趋势，这是数据新闻中的重中之重，是粘合读者的方法措施，也是大众的需求所在。

2.4 专业人才缺乏，团队建设滞后

传统的新闻报道，某个记者或者自由撰稿人就可以独立完成。但是优秀的数据新闻作品一般需要团队合作，因为生产过程复杂、耗时巨大，包括新闻选题、数据搜集整理、数据分析和数据可视化等协同合作。目前，国内的数据新闻起步相对滞后，团队大多由传统新闻的编辑成员组成，不可避免地导致团队缺少对数据新闻学习的过程，整体的数据新闻制作专业化程度较低。由于数据挖掘、分析能力欠缺，致使其在选题上无法获取更多有价值的线索，在新闻呈现方面更是大同小异，同质化较为严重。好的数据新闻作品，需要生产者具备高超的数据素养，具备敏锐的数据感觉和分析能力，在海量的数据库信息中，能够分辨出有价值的数据，挖掘背后的故事，拟成有价值的新闻选题，能够基于数据库和数据分析建立数据模型，对同类信息进行监测，从而写出预测性的新闻报道，进一步提升新闻的价值。

2.5 数据可视化工具普及度不高

因为数据分析、数据可视化的专业性，数据可视化工具在媒体行业的普及程度并不高。像 Tableau、Plotly、Rggplot2、Raw、Infogram 等可视化工具，对于传统媒体从业人员来说已经足够陌生，更不必说像 R 语言、Python 等更加底层专业的可视化编程语言。因为数据可视化工具普及度不高，现有的可视化新闻作品，很多都是通过记者编辑辅以设计师来完成，如果涉及更深层次的数据分析与更加炫酷的数据呈现，可能还需辅以更多专业性人才，如程序员、数据分析师等，这对于传统媒体来说，无疑是一笔不小的成本。

3. 数据新闻可视化的发展策略

针对数据新闻可视化存在的种种问题，传统媒体应该如何去解决，将数据新闻可视化做得更好呢？

3.1 发挥媒体优势，扩展数据来源

数据是新闻制作的关键，获取海量、有效的数据，创建高质量的数据库，是促进数据新闻发展的第一步。

3.1.1 推动数据开放与共享

相关部门与企业掌握着大量的数据，税务部门掌握着纳税数据，交通管理部门掌握着交通数据，零售企业掌握用户购买数据，互联网企业掌握用户上网行为数据，制造企业生产过程中也产生了大量的数据，但这些数据目前大多数都被保护着，没有得到合理的开放与共享。站在媒体的角度上，可以从“数据开放自助、数据质量管理、数据合法保护、数据保存和数据共享利用”等方面推动数据开放进程。借鉴国外优秀媒体经验，推动公开信息，建设政务数据库，梳理政治敏感，提高数据处理能力，稀释数据，防止数据垄断。同时管理好企业的重要数据，打造数据交易流通的平台，开设

监管部门，制定网络法律规范，健全把关行为监督机制，确保信息的合法、安全使用，为国内数据新闻的发展提供更加规范、全面的数据库。

3.1.2 整合物联网数据，融通社交媒体资源

丰富数据来源，充分挖掘利用社交媒体资源，并有效地整合物联网数据，增强数据新闻的传播效果。譬如遍布城市的监控摄像头，每天收集大量的数据，如果经过有效的分析及运营，必定能够产生巨大的价值。以两微一端为代表的社交网络媒体风头正盛，其反应速度往往比传统媒体更加迅速，内容更丰富，也能收获更多读者的反馈。依托社交网络的分享优势，收集网络用户的数据，精细地提供个性化服务，从国家和社会的层面，推动到个人层面，共享数据红利。

3. 2 利用多媒体技术，提升可视化交互水平

利用不同平台的优势，制作契合该平台的新闻作品，找到每个媒介最适宜的呈现方式。从视觉形式看，可视化的类型可以拓展为数据地图、时间轴、信息云、人物关系图等。图形的运用，延伸和补充了信息图的内涵。从媒体形式看，可借助音 / 视频、三维技术、虚拟现实技术和人脸识别等多媒体技术，提高数据新闻的丰富度与审美趣味，让原本的“平面化”变得更加“立体化”。不仅能满足视觉上的观感，还能增加听觉和触觉的代入感，增强新闻的临场感与传播力，重塑阅读体验。

在可视化过程中增加反馈渠道，提升读者信息反馈价值。一方面是人与信息的交流，读者的评论意见可以帮助新闻制作方更客观地看待新闻内容和信息传播。另一方面，人与人之间的点赞回复，促成更为活跃的信息流通，增强有效信息的传播。与此同时，观点的碰撞能够帮助读者消除信息的不确定性，有助于更理性地理解新闻信息。

3. 3 提高媒体从业人员的数据素养

3.3.1 与高校合作增设交叉学科课程，储备人才资源

目前，新闻传播、计算机技术以及美术设计等专业在高校并不少见，但在数据新闻可视化领域却还存在一定的短板。因此，将传统媒体的需求与高校教育体系相结合，适当增设交叉学科的课程，可以促进多学科背景人才的培养，为数据新闻可视化提供人才储备。

在课程设置上，将新闻传播学和经济学、计算机、法学、艺术设计等专业横向联合，不仅可以充分利用高校优秀的教育资源，而且有利于培养多向发展的综合性人才。辅以媒体实践，加倍注重培养学生对数据的敏感度，提高甄别数据真伪的能力，用数据挖掘分析和数据可视化技术全面武装自身，探索数据背后的深层含义，为将来从事数据新闻报道奠定坚实的基础。

3.3.2 转变思维能力，引进专业人才

可视化的初衷是为了直观呈现复杂数据背后的信息以及趋势，不应该为了可视化而可视化，过于花哨的设计和装饰会掩盖数据新闻的内容。因此，信息图表的设计应当以简洁和清晰明了为主要目标，明确数据信息，突出重点。除此之外，新闻从业人员需要敢于打破思维禁锢，接纳更新鲜的事物。例如让机器参与日常的工作任务，把自己从基础、繁重的工作任务中解放出来，去从事更具智慧性和创造性的工作。

注重数据新闻人才培养，引进对新闻和数据敏感、具备分析挖掘和整合数据能力以及擅长数据可视化技术等方面的人才，组建专业化的数据新闻可视化制作团队。新闻机构需要设置相应的岗位培训课程，提升新闻从业人员的职业素养，帮助他们加快理解当下的时代背景和媒体环境，熟悉数据可视化工具与生产流程，从而提升数据甄别、分析和挖掘能力，形成自身的一套洞察、认知和行为逻辑，更好地为人们提供信息服务。

3. 4 深度挖掘数据价值

3.4.1 提升选题价值，对事件进行多维度交叉对比

对于新闻题目的选择，是传播学中“把关人”身份的体现。通过科学有效的技术将数据进行分析解读，从全局着眼，增加选题的重要性，发掘数据、事件和人之间的内在联系，帮助读者发现事件与个人的联系以及对人的影响，找到自身与新闻的连接点，从而提升新闻与读者的接近性。

在对事件进行分析或是解读时，从不同维度搜集信息，通过数据的叠加和对比，揭示数据背后的联系，尽可能拼凑事件的全貌。利用不同维度对数据进行分析，可以发现表面上关联度不大的数据之间深藏的底层关系。

3.4.2 推广数据众包模式，媒体与用户参与其中

依靠社会群体的力量，尽可能地寻找新闻事件发生、发展和结果等碎片，激发大众参与新闻报道的热情，产生“1+1>2”的效果，进而推动数据新闻制作的创新发展。如果仅凭新闻记者的一己之力，结果很有可能不尽如人意，且难以满足读者对新闻资讯的渴求和置身其中的快感。采用众包的方式收集数据，邀请用户主动参与和分享，上传个人发现的数据等，通过开放实时评论平台、电子问卷等方式获取信息，既保证了数据的来源，又降低了收集数据的成本。因此，数据众包是数据新闻可视化探索尝试新发展的路径之一。

4. 数据新闻可视化的发展趋势

随着智能设备的普及，数据产生的渠道越来越多样化，新闻呈现的方式越来越多样化，数据分析、新闻可视化必定成为传媒行业未来发展的主流趋势。特别是移动智能终端的普及，将推动数据新闻的不断发展与成熟，移动端数据新闻作品会成为新闻播报的常态形式。大数据、人工智能、云计算等技术的成熟，或许会颠覆传统媒体的生产方式，也将为数据新闻的可视化提供更加便捷的生产工具。

数据对人们行为的分析越来越准确，为读者量身定制所需的行动方案提供了可能。在大数据的基础上，制作出具有普遍性的数据新闻，再以用户为中心，将信息可视化定制式地融合到用户的日常任务中，设计信息可视化环境，促进用户的合作与共享。这不仅是对数据的再次开发利用，也是融入新媒体舆论场，巩固并扩大新闻影响力的一个重要手段。

数据新闻的制作是一个综合的过程，需要多方面的人才通力合作。无论是新闻编辑，还是设计工作者，数据处理能力的培养和提升均迫在眉睫。一言以蔽之，关于可视化新闻的探索刚刚开始。随着经济的发展和大众媒介素养不断提升，可以预见数据新闻在未来会有更加广阔的发展前景，相信在

（下转第 125 页）